DOCUMENT RESUME

ED 451 572 CS 510 542

AUTHOR Sunal, Cynthia Szymanski; Sunal, Dennis W.; Tirri, Kirsi

TITLE Using Evidence in Scientific Reasoning: Exploring Characteristics of Middle School Students' Argumentation.

PUB DATE 2001-04-00

NOTE 35p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).

PUB TYPE Numerical/Quantitative Data (110) -- Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Communication Research; Communication Skills; Discussion (Teaching Technique); *Group Dynamics; Instructional Effectiveness; *Middle School Students; Middle Schools; *Persuasive Discourse; Scientific Methodology; *Small Group Instruction; *Thinking Skills

IDENTIFIERS Evidence

ABSTRACT

This study considered the ability of small groups of students aged 13-15 years to structure and argue an evidentiary case in which the limitations of the evidence are recognized. The findings of this study indicate that these student groups had developed some of the characteristics of argumentation and group process skills used in building and arguing an evidentiary case. However, these students and their student groups need guidance if they are to further develop their use of evidence and recognition of its limitations in argumentation. The findings further indicate that it was possible to categorize students' argumentation in small group discussion based on their ability to use evidence and identify its limitations. (Contains 23 references and 13 tables of data.) (Author/RS)

Reproductions supplied by EDRS are the best that can be made from the original document.

ED 451 572

Using Evidence in Scientific Reasoning: Exploring Characteristics of Middle School Students' Argumentation

Cynthia Szymanski Sunal

The University of Alabama
Tuscaloosa, Alabama

Dennis W. Sunal

The University of Alabama
Tuscaloosa, Alabama

Kirsi Tirri

University of Helsinki, Finland

Paper presented at the annual meeting of the American Educational Research Association, Seattle, Washington, April, 2001

RUNNING HEAD: Argumentation

Address Correspondence to:
Cynthia Szymanski Sunal
The University of Alabama
P.O. Box 870232
Tuscaloosa, Alabama 35487-0231
FAX: 205-348-9863
phone: 205-348-7010
email: cvsunal@bamaed.ua.edu

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
C. Szymanski Sunal
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

CS 510 542

Abstract

This study considered the ability of small groups of students aged 13-15 years to structure and argue an evidentiary case in which the limitations of the evidence are recognized. The findings of this study indicate that these student groups had developed some of the characteristics of argumentation and group process skills used in building and arguing ar /identiary. However, these students and their student groups need guidance if they are to fur' develop their use of evidence and recognition of its limitations in argumentation. The findings further indicate that it was possible to categorize students' argumentation in small group discussion based on their ability to use evidence and identify its limitations.

Using Evidence in Scientific Reasoning: Exploring Characteristics of Middle School Students' Argumentation

Argumentation of evidence is a part of every discipline. Yet, students are not well prepared to consider evidence and use it to build a case arguing for a particular response to a question. Recently, scholars in fields ranging from history to physics have called for greater attention in the curriculum to the use of evidence and its argumentation (Yerrick, 2000; Driver, Leach, Millar, and Scott, 1996; Greene, 1994). A lack of attention to building a case for a response to a question can lead to students' unquestioning acceptance of material delivered in a lecture or through a book (Kuhn, 1993; Driver et al). It is easier to accept or reject someone else's viewpoint or stance outright than to consider the evidence for and against it, the limitations of that evidence, and how that evidence was interpreted and argued. Indeed, many students seem to have little understanding that a case is built slowly, using and arguing evidence (Kuhn). They do not understand that the process of building a case for a specific response to a question is the lifeblood of a discipline. The concept of a case used throughout this paper is that of a narrative framework centered upon an event or events for which interactions are analyzed and conclusions are reached. The analysis incorporated within a case involves an understanding of context, chronology, and the interpretation of evidence.

Students need to develop the skills necessary to ask questions and gather and analyze evidence. Then, they must interpret that evidence, building a case they argue with their peers (Sunal & Sunal, 1999). Throughout this process, a key component is the estimation and comprehension of the limitations of the evidence. This study addressed middle school students' use of evidence in scientific reasoning and argumentation in the context of a unit on nuclear fusion as an energy source. The specific focus of the study was on the students' understanding of the limitations of evidence and interpretation of the meaning of that evidence in terms of the arguments they constructed.

Collaborative reflective learning occurs during argumentation. The student is challenged, encountering: 1) differing interpretations, 2) the effects of the context in which an interpretation is constructed, 3) the need to be aware of, and judge, the limits of the evidence, and 4) the need to be systematic in the presentation of the evidence (Sunal & Sunal, 1999). Anomalous data can set up a challenge to students' thinking. Chinn and Brewer (1998) reported that students' use of anomalous data takes eight forms: ignoring the data; rejecting it outright; excluding it by declaring it irrelevant; holding it in abeyance by deciding it is insufficient or has too many

uncertainties associated with it; reinterpreting it by arguing that the causal explanation is significantly different from that proffered by the scientist; modifying theories peripherally by arguing the data's effects are minor rather than major; and, expressing uncertainty about the data itself. In only 8 of 168 cases in their study did students modify their views as a consequence of evidence contradictory to their previously held beliefs. Making students aware of how both they, and scientists, respond to contradictory claims provides important insights into the social processes internal to scientific argument. Contradictory claims can highlight limitations of the evidence in use. If students do not easily modify their views when encountering contradictory evidence, they may have difficulty identifying limitations in the evidence they are using.

Argumentation is a web of skills used to organize evidence into a framework supporting a hypothesis, viewpoint or stance. As a question is asked, evidence is gathered, analyzed, and evaluated to build a case arguing for a particular resolution of the question. The evidence is subjected to analysis using a set of criteria that may or may not have been predetermined. Often, the criteria are established before the evidence is examined. But, the evidence may first be examined without criteria in order to try to find patterns or otherwise establish criteria by which to analyze and evaluate it. When criteria are established, they take a form that enables others to examine the criteria, evidence, analysis, and evaluation conducted using the criteria. Students' development of the ability to structure, argue, and restructure a case built on evidence is important to their development as scientifically literate individuals. To facilitate such development it is necessary to examine how students interpret what the available evidence means and what limitations the evidence has.

Reasoning and Argumentation

The use of evidence to build and argue a case occurs is procedural knowledge. Hypothesizing also is a procedural knowledge skill. It is related to the use of evidence to build and argue a case as a possible answer to the question around which an investigation is based. Scientists have extensively studied the formulation of hypotheses. In science, an investigation often has been perceived as beginning with a hypothesis to be tested rather than leading to a hypothesis (Popper, 1934; Kuhn, 1962). Science is seen as making progress through the replacement of hypotheses by newer ones with greater empirical content; that is, they account for a larger number of observations. This progress occurs as current hypotheses are recognized as deficient through experimentation. The key criterion a hypothesis must meet is that it has testable consequences that could lead, in principle, to its falsification (Popper). While Popper focuses on the refutation aspect of scientific investigation, Kuhn talks about how hypotheses frame the experimental design and indicate expected observations thereby influencing data collection and the possibility of bias. Both Popper and Kuhn consider hypotheses and their testing as central to scientific investigation.

Six clusters of ideas related to the construction of an evidentiary case and reasoning have been identified (Leinhardt, Stainton, Virji, & Ordoff, 1992; Sunal & Sunal, 1999). First, the investigation implies a purpose or motivation. The motivating assumption underlying the construction of a case from evidence is that the case exists to help us understand both what was and also what is (Levstik & Pappas, 1987). Second, the investigation leads to the construction of a case with internal coherence (Leinhardt, Beck & Stainton, 1994). Third, a case is exhaustive, analyzing and synthesizing all evidence that can be found and all evidence that might contradict the case (Wineburg & Fournier, 1994). Fourth, the case is the answer constructed in response to critical questions selected through the use of theory or as part of hypothesis development. Fifth,

the evidence from non replicable events, such as a fossil layer on a cliff cut away by a stream bed, is interpreted in terms of the context of the original times when events occurred, the context of the present time seeking to look at that evidence, and the implications of which evidence has survived over time. This is contextual, layered interpretation. Sixth, argument or debate of other competing interpretations occurs and includes organization of all relevant documentation.

An Argumentation Model for Student Instruction

Many cognitive processing skills are used to structure the case that is argued. A seven-step process can be described beginning with a question or problem, leading to hypothesis generation, then to evidence gathering, next to construction of an argument, on to further hypothesis development based on the reflection involved in argument construction, next to hypothesis testing, and finally to related new questions or problems (Sunal & Sunal, 1999). A student or a researcher can, and often does, enter the process at any step. The process is a recursive one since a new question, problem, and hypothesis lead to evidence gathering, the construction of an argument, debate of the argument, and so on. These are basic components of a process used to help students develop argumentation. As such, it is a model for curriculum and instruction aimed at using cognitive processing skills in science. The model is presented as a means of suggesting starting points and components.

The research base in science that has examined the acquisition of argument construction and its sub components has found a broad lack of ability among students and adults. However, students involved in experimental inquiry-based science in elementary and middle school may have proficiency in the main components of the model. Rosalind Driver and Paul Newton (1997) noted that "Science teaching has paid little attention to argument and this has led to important shortcomings (p. 1)." In particular, they stated that the lack of attention to argumentation in science has led to a false impression of ". . . science as the unproblematic collation of facts about the world (p. 1)." When many people encounter controversies in science these are puzzling since science is viewed as a set of facts. This context for science instruction leads to students' lack of ". . . ability to argue scientifically through the kinds of socio-scientific issues that they are increasingly having to face in their everyday lives (p. 1)" (see also Solomon, 1991; Norris & Phillips, 1994).

Associated research has examined the effects of cognitive intervention in formal cognition skills. In the United Kingdom, Shayer and Adey (1990) examined cognitive intervention in science during middle school with students aged 11 and 12. The intervention focused on variables and multiple variables, proportionality, compensation, probability, correlation, classification, equilibrium, and formal models. Nontraditional, cooperative group oriented, inquiry-based training methods were used. By the end of the intervention period, the experimental group had achieved a significantly greater gain in levels of cognitive processing skills than had the control group. Shayer and Adey (1993) reported long-term effects four years after the initiation of their intervention in terms of increased performance by experimental students on the relatively objective norms of the science, mathematics, and English measures administered as part of the United Kingdom's national examination system, the General Certificate of Secondary Education (GCSE). While the intervention reported large and permanent effects on students' achievement of cognitive processing skills, the repertoire of cognitive skills needed to use evidentiary data in building a case for its interpretation and in arguing that case were not part of Shayer and Adey's research effort.

An examination of students' consideration of science-technology-society issues could

give insights into their use of evidence in argumentation of science ideas. Such investigation could use topics found in middle school science programs that incorporate evident science-related social issues such as "energy" (American Association for the Advancement of Science, 1993; National Research Council, 1994).

Analysis of students' evidence gathering involves three components. First students examine the existing evidence base. Second, they should have opportunities to work toward exhaustive evidence gathering. As part of the process they will need to be helped to understand that exhaustive evidence gathering is not likely to be possible because of the limitations of their own maturity and skills, time, money etc. However, the understanding that exhaustive evidence gathering is a goal toward which scientists strive is important. Finally, students must come to understand the limits of the evidence available to them. They must comprehend that not only are they probably unable to carry out exhaustive evidence gathering, but that the evidence they do have has limitations. Such critical analysis is difficult. Group processes, when used within a classroom, have the potential to challenge students to recognize the limitations of evidence since learning is socially situated within the classroom and context is an important part of an argument (Lee, Dickinson, & Ashby, 1994; Driver & Newton, 1997). Group processes that may encourage the consideration of the limitations of available evidence include whether students check the coordination of evidence and claims, encourage the presentation of different ideas, and attempt to coordinate different perspectives (Driver & Newton).

This study involved students in the interpretation of evidence and consideration of its limitations as they explored a potential energy source, nuclear fusion and related science-technology-society issues. Students were involved in: laboratory experiments on fundamental energy concepts in electricity and magnetism, the use of briefing papers, a panel presentation, discussion of the quality of evidence available and of the limitations of the evidence; and finally, in voting by ballot as a member of a decision-making body considering whether nuclear fusion is the best and safest source of energy for our future needs.

The Study

The purpose of the study was to determine whether students' argumentation in small group discussion can be categorized based on their ability to use evidence and identify its limitations. Framed questions guiding the study were: What are the characteristics of argumentation occurring in small groups? What are the characteristics of students' understanding of the limitations of evidence? What interpretations do students make of the evidence they have in terms of the arguments they are making? and What group processes are undertaken by students when involved in interpreting evidence and understanding its limitations?

Procedures

This study involved students in seven classes at two sites. All of the classes worked in cooperative learning groups at least twice weekly in their science class. At school A, participating students were from two intact ninth grade earth and physical science classes of 25 students each taught by the same teacher in a high school enrolling students from region that was rural and small town in composition. Table one describes data by class for each school and class including size, number of males and females, age range, and mean age. The mean age for participating students from school A was 15.0. Students in both classes were considered by the school system to be achieving at an average or above average level based on grade point average

and scores of the state mandated Stanford 9 Achievement Test (SAT). Scores on the SAT categorized this school as average within the state scoring at the 50^{th} – 60^{th} percentile. These students were 70% European American and 30% African American.

At school B, five sections averaging 25 students of eighth grade earth and physical science students taught by the same teacher in a middle school participated. These 100 students were all African American and lived in a small city in the southeast with 150,000 residents having a mixed financial base including industries such as a steel plant, a new textile plant, and a new auto assembly plant, and a large university with several thousand employees. The city was adjacent to the county in which school A students attended school. Table one gives data by class. The mean age of school B's participating students was 14.2, so these students' mean age was eight months younger than that of students at school A. Class three was considered by school administration and the school system's special education services department to contain students mostly achieving at a gifted level. Students in this class ranged from 13.7- 14.1 in age with an average of 13.9. Class five contained 24 students including 11 females and 13 males and was considered to be achieving at a below-average level. Students in this class had the oldest mean age at School B. Students in classes four, six, and seven were considered by the school system to be achieving at an average or above average level based on grade point average and scores on the SAT. Scores on the SAT categorized this middle school as low performing within the state averaging in the 30^{th}-40^{th} percentile range.

Table 1 about here

The two teachers of these participants had volunteered for the study because of an interest in fostering argumentation skills among the students. Observations over a period of two months were conducted in the teachers' classrooms to establish the content and instructional strategies involved in the nuclear fusion unit. The students had been working in cooperative groups chosen through a stratified random selection procedure based on gender at the beginning of the school year. The target group was four students including two males and two females. However, because of unequal male-female balances in the classes, eleven of the 38 groups had five students. In each group of five students at least two represented each gender. These groups of five students occurred as follows: class one – one of five groups; class two - five of five groups; class three – one of six groups; class four – two of six groups; class five – one of five groups; class six – one of five groups; and class seven – one of six groups. Because of their experience throughout the school year, these students were familiar with group processes and had opportunities in taking responsibility for group tasks.

In this study, students worked in cooperative groups conducting experiments, teaching other group members about the content of briefing papers provided for them, and preparing for whole group discussions. The group leader used prompts to guide the small group discussions including: "What are the limitations of the evidence you have about nuclear fusion?" "Where do you need more information, or different information, in order to make a decision about a question and give an answer you think is supported by evidence?" "Why can't you get all the information, or evidence, you think you need to give you a good answer?" "Where could you get more information, or evidence?" "When you think about where you need more information, or different information, what does this tell you about how sure you can be of your answers to the questions we talked about?"

Transcripts were made of small group discussion. These were analyzed with a protocol that examined whether the students compared hypotheses put forward in the light of the evidence

that was available so that they could move to justify their positions. The protocol also examined whether the teacher made sure that different ideas were listened to and understood. Evaluation criteria in the protocol were established for three areas: 1) the characteristics and quality of small group arguments that indicate students' understanding of the limitations of the evidence available, 2) the characteristics of students' interpretation of the meaning of the evidence available, and 3) the structure and use of evidence in the case students built for argumentation. Students were involved in a unit on nuclear fusion designed in a learning cycle format. First, they were involved in an exploration whose objective was to involve them in identifying possible evidence than can be used in the solution of the problem, "What source of energy is the best and safest for our future needs?" "What evidence supports your response?" They discussed the question in small groups, reported their conclusions to the whole group, and summarized their ideas on chart paper posted for all to see.

The invention that followed utilized an initial whole class discussion, three experimental investigations, a brief reading introducing nuclear fusion considered by all, and experts' briefings. The first experimental investigation focused on working with the effects of a closed electric circuit on a magnetic compass. The second investigation involved students in experimenting with looped closed electric circuits and their effects on a magnetic compass. In the third investigation, students built an electromagnet with a battery, wire, and a large iron nail. They tested the electromagnet's ability to attract objects from a set they were given. They also investigated the effects of varying the strength of a magnetic field as the number of loops of wire around the nail was increased and decreased. A whole group discussion reviewed the results of the experimental investigations and considered their relationship to the brief reading introducing nuclear fusion.

Following the experimental investigations and whole group discussion, each member of a group was given a set of expert briefing materials and was responsible for teaching the material in his or her expert briefing to other group members. The expert briefings considered the problem of containment; problems related to the conditions for nuclear fusion including temperature, plasma density, and plasma containment; long-term energy needs in the world and means by which they might be addressed; and advantages and disadvantages of nuclear fusion as an energy source.

The group manager led two discussions of each expert's materials using a set of questions provided by the teacher. The first discussion involved group members in sharing information from the expert briefings. During the second discussion, group members considered issues related to nuclear fusion. The discussion questions encouraged students, and their groups, to examine how well they were able to consider the limits of the evidence they had gathered and what their evidence meant. The teacher concluded this part of the learning cycle by asking the question, "What is evidence?" The students developed a whole group definition for "evidence" based on their discussion thus far.

The expansion portion of the learning cycle involved students in a panel discussion followed by a vote on the question "Is nuclear fusion the best and safest source of energy for our future needs?" Students from each group first wrote out what they thought the limitations of their evidence were, and what they thought their evidence meant. The group's reporter then served as a representative on a panel considering the world's energy problems. Each reporter represented a group of experts currently doing research on nuclear fusion. The following prompts were used by teachers. "What are the limitations of the evidence about nuclear fusion that you have?" "Where do you need more information, or different information, in order to make a decision about a question and give an answer that you think is supported by evidence?" "Why can't you get all the

information, or evidence, you think you need to give a good answer?" "Where could you get more information, or evidence?" "When you think about where you need more information, or different information, what does this tell you about how sure you can be of your answers to the questions we talked about?" Students then voted on the question, "Is nuclear fusion the best and safest source of energy for our future needs?" Their ballots first asked them to vote yes or no on the question. Then, they were asked to give their reasons for the answer checked using the questions, "What evidence supports your answer?" and "How sure are you about your answer?" "Why?

A small audiotape recorder was placed on the floor near each group. During its' first usage, the groups were told about the recording device and allowed to inspect and play with it for three minutes. They were told they could cover up the device during the discussion should they wish to do so. On a few occasions groups chose to do so for up to four minutes. The students' study of nuclear fusion took ten days of 50 minute periods of class time.

Analysis

Transcripts of audiotaped small group discussion during the invention portion of the learning cycle were analyzed using three schemes. The first incorporated portions of a provisional set of hierarchical criteria devised by Driver and Newton (1997) that drew upon the characteristics of argument proposed by Tolumin (1958). The criteria were used to evaluate small group argumentation structure. These characteristics and the level at which they are scored included the following: 1). single claim with no reasons, level 0; 2). competing claims with no reasons, level 0; 3). single claim with reason(s), level 1; 4). competing claims with reasons and qualifiers, level 3; 5). claim(s) with reasons responded to by rebuttal, level 3; and 6). making judgment(s) integrating different arguments, level 4.

The second analysis scheme, Identification of Evidence and its Limitations in Argumentation, used criteria to evaluate the ability of small groups to build a case for argumentation incorporating the use of evidence and recognizing its limitations. The criteria were scored as a "one" indicating presence of the criterion or as a "zero" indicating absence, for a possible maximum score of ten. Three subgroups of criteria were scored. The criteria were as follows.

Subgroup A: Examination of an existing evidence base
Gathering of evidence beyond that provided by teacher
Exhaustive evidence gathering
Subtotal (0-3)
Subgroup B: Identification of one limitation of evidence available
Identification of more than one limitation of evidence available
Consideration of the context of the times
Subtotal (0-3)
Subgroup C: Citation of evidence
Linkage of two pieces of evidence in an argument
Organization of evidence into a meaningful structure
Consideration of counter evidence
Subtotal (0-4)
Total score (0-10)

Subgroup A considers students acquisition of evidence from usage of an existing evidence base such as that provided by a teacher to attempts to carry out exhaustive evidence gathering.

Subgroup B considers students identification of limitations. The context of the times involves comprehending how social influences and perspectives related to the time period under consideration may affect scientific research and findings or determine what hypotheses are studied, what instrumentation is available, and how findings are applied. Students may be able to operate at any level of subgroup B in combination with any level of Subgroup A. Thus, they may utilize only the evidence base provided by the teacher but may identify multiple limitations of that evidence base. Subgroup C considers students building of a case to argue the evidence. Again, subgroup C is not necessarily related to subgroups A and B. A student group may recognize limitations of evidence but be unable to organize it into a meaningful structure or consider counter evidence. An assumption underlying the scheme of this instrument is that students who, in subgroup A, pursue additional evidence gathering may identify multiple limitations of the evidence (subgroup B) and may consider counter evidence (subgroup C) because the criteria are structured within in each subgroup in terms of their increasing abstractness and complexity.

The third scheme was derived from Driver and Newton (1997) and evaluated the group processes students used. These were rated as a "one" if present or as a "zero" if absent with a possible maximum score of seven. They included: check information and evidence; question each others' reasons; check coordination of evidence and claims; build on each others' arguments; monitor involvement of group members; encourage presentation of different ideas; distinguish between scientific claims and those based on other types of knowledge.

Findings

The findings of the study are reported in three parts, consistent with each of the analysis schemes used. Findings for each analysis scheme are presented for the small groups. Then, the groups are considered by school site. Since school site B had five classes including three designated by the school and school system as average to above average in performance, one designated as gifted, and one designated as low performing, the findings are then considered by class designation within school B. Findings on he characteristics of the arguments made in the small groups are examined first. Next, findings on students' use of evidence and identification of its limitations are presented. Finally, students use of group processes during discussion is reported.

Tables 2, 3, 4 about here

Characteristics of the Arguments

Using the analysis scheme derived from Tolumin (1958) by Driver and Newton (1997) arguments were examined in a progression of four levels demonstrating increasing complexity of thought beginning with making a single claim with no reasons given and ending with making judgment(s) integrating different arguments (see Tables 2, 4). These students mostly fell in a range of three characteristics of argument at two levels. Eight groups made level 1 arguments involving a single claim with reasons given. Twenty-seven groups demonstrated level 3 argumentation characteristics with eleven making competing claims with reason(s) given and qualifier(s), and sixteen making claim(s) with reason(s) responded to by rebuttal. One group

demonstrated argumentation at level 4 involving making judgment(s) integrating different arguments while one group each demonstrated level 0 argumentation, making a single claim with no reasons given or making competing claims with no reasons given.

Argumentation by student groups at school A showed level 3 and 4 characteristics with 50% of the groups making claim(s) with reason(s) given and qualifier(s) (see Table 3). Another 40% made competing claims with reason(s) given and qualifier(s) while 10% made judgment(s) integrating different arguments. At school B, 50% of the groups also made claim(s) with reason(s) given and qualifier(s) but no group's argumentation demonstrated characteristics found at level 4 of argumentation. Competing claims with reason(s) given and qualifier(s) were made by 25% of the small groups at school B while 29% made a single claim with reason(s) given and 8% of the groups demonstrated argumentation characteristics at level 0.

Differences were found among the five classes at school B when considered by class designation of average to above average in performance (three classes), gifted (one class), or low-performing (one class) (Table 3). Student groups designated as average to above average demonstrated characteristics three, four, and five, found at levels 1 and 3. Most (47%) demonstrated level 3 characteristic five, claim(s) with reason(s) given and qualifier(s) while 35% showed the level 1 characteristic three and another 24% showed level 3 characteristic 5. The class designated as gifted demonstrated argumentation at level 3 with 67% displaying characteristic 5 and 33% characteristic 4. The class designated as low performing demonstrated characteristics spread across levels 0 to 3. Two groups demonstrated level 1 characteristic three while one group demonstrated level 3 characteristic four and one group each demonstrated level 0 characteristics one and two.

Tables 5, 6, 7, 8, 9, 10, 11, 12 about here

Use of Evidence and Identification of its Limitations

The use of evidence by small groups and their identification of its limitations was examined using criteria in three subgroups. Subgroup A considered whether the small group had: 1). examined an existing evidence base, 2). gathered evidence beyond that provided by the teacher, and 3). carried out exhaustive evidence gathering (see Tables 5 and 6). Subgroup B considered whether the small group had: 4). identified one limitation of the evidence available, 5). identified more than one limitation of the evidence available, and 6). considered the context of the times (see Tables 7 and 8). Subgroup C considered whether the small group had: 7). cited evidence, 8). linked two pieces of evidence in an argument, 9). organized evidence into a meaningful structure, and 10). considered counter evidence (see Tables 9 and 10).

Subgroup A

All but two groups (95%) examined the existing evidence base provided to them by their teacher. Many (84%) gathered evidence beyond that provided by the teacher. However, just seven groups (18%) carried out efforts at exhaustive evidence gathering (see Tables 5 and 6). These groups searched the Internet, contacted researchers carrying out research on nuclear fusion at Auburn University, Alabama, talked with local power company officials about energy production, and/or carried out their own further experimentation with electromagnets. Those other groups carrying out some additional evidence gathering that was not at a level evaluated as

exhaustive, searched the Internet.

As Table 11 shows, at school A 100% of student groups examined the existing evidence base while 93% did so at school B. Gathering of evidence beyond that provided by the teacher occurred in 90% of the groups at school A and 82% of the groups at school B. Exhaustive evidence gathering was noted in 20% of school A's groups and 18% of school B's groups.

In school B, 92% of the groups in classes designated as average to above average examined the existing evidence base while 100% of those in the class designated as gifted and 67% of the low performing class did so. Gathering of evidence beyond the provided by the teacher occurred in 94% of the groups in the average to above average classes, 100% of the gifted class, and 17% of the low performing class. Exhaustive evidence gathering was noted in 24% of the average to above average classes, 17% of the gifted class, and not at all in the low performing class.

Subgroup B

All but one group of students, in school B, identified one limitation of the evidence available (see Table 7). Many groups (74%) identified more than one limitation of the available evidence. The context of the times in which the issues were set was considered by 74% of the groups. As shown in Table 11, 90% of school A's groups identified more than one limitation of the evidence available to them from the teacher while 68% of school B's groups did so. The context of the times was considered by 80% of school A's groups and 50% of school B's groups. Within school B, 94 % of groups designated as average to above average, 100% of groups designated as gifted, and 100% of groups designated as low performing identified one limitation of the available evidence. The identification of more than one limitation of the available evidence occurred in 65% of the average to above average groups, 83% of the gifted class's groups, and 50% of the low performing class's groups. Consideration of the context of the times occurred among 55% of the groups including 80% at school A and 50% at school B. Within school B, the context of the times was considered by 53% of the above average to average classes groups, 100% of the gifted class, and 17% of the low performing class's groups.

Subgroup C

The use of evidence in group argumentation is considered by the criteria found in Subgroup C (see Table 9). Evidence was cited by all groups. Two pieces of evidence were linked in making an argument by 95% of the groups. The evidence was organized into a meaningful structure by 66% of the groups. Counter evidence was considered by 39% of the groups. In school A all of the groups cited evidence while all but two groups (93%) did so in school B. The evidence cited was organized into a meaningful structure by 60% of school A's groups and 68% of school B's groups. Counter evidence was considered during argumentation by half of school A's groups and 21% of school B's groups. When considered by class designation, 88% of the groups in school B's average to above average classes cited evidence in their argumentation as did 100% of the groups in the gifted class and 83% of groups in the low performing class. Two pieces of evidence were linked in their arguments by most (94%) of the groups in the average to above average classes, all of the groups in the gifted class, and 67% of the groups in the low performing class. The evidence was organized into a meaningful structure by 71% of the groups in the average to above average classes in school B, 83% of the groups in the gifted class, and 33% of the groups in the low performing class. Counter evidence was considered by 18% of the

average to above average classes groups, 50% of the gifted class's groups, and none of the low performing class's groups.

Use of Group Processes in Argumentation

Seven processes found by Driver and Newton (1997) to be facilitative of the involvement of all members in small group argumentation were considered (see Tables 13, 14, and 15. All but two groups checked information and evidence (95%). Most groups (89%) questioned each other's reasons during argumentation. Many of the groups (71%) checked the coordination of evidence and claims made during discussion. Fewer (53%) built on each others' arguments. Monitoring to insure the involvement of each group member occurred in 84% of the groups. The presentation of different ideas by group members was encouraged in 68% of the groups. However, only 39% of the groups tried to distinguish between scientific claims and those based on other types of knowledge.

Information and evidence were checked by all groups in school A and all but two groups in school B (93%) as indicated in Table 14. Students in all of school A's groups and 86% of school B's groups questioned each other's reasons. The coordination of evidence was checked by 80% of school A's groups and 68% of school B's groups. As they carried out their discussions, 60% of school A's groups and 50% of school B's groups built on each others' argumentation. All of school's A's groups and 79% of school B's groups monitored the involvement of individual group members. Individuals within the group were encouraged to present different ideas in 70% of school A's groups and 68% of school B's groups. Half of school A's groups and 36% of school B's groups distinguished between scientific claims and those based on other types of knowledge.

School B group discussions were considered by class designation. It was found that all of the average to above average classes and the gifted class groups checked information and evidence but just half of the low performing class groups did so. Each other's reasons were questioned by 94% of the groups in the average to above average classes, 100% of the gifted class groups, and 33% of the low performing class groups. The coordination of evidence and claims was checked in 76% of the average to above average classes' groups, 67% of gifted class groups, and 33% of low performing class groups. Students built on each others' arguments less often: 59% in average to above average classes' groups, 50% in gifted class groups, and 17% in low performing class groups. More of the groups monitored the involvement of group members: 76% in average to above average classes, 83% in the gifted class, and 67% in the low performing class. Encouragement of the presentation of different ideas was found among groups in 71% of average to above average classes, 67% of the gifted class, and 50% of the low performing class. The least used group process was distinguishing between scientific claims and those based on other types of knowledge. This was found among 47% of the above average to average classes' groups, 17% of the groups in the gifted class, and 17% of the groups in the low performing class.

Conclusions and Implications

The study investigated students' use of evidence and understanding of its limitations in building and arguing a case on an energy problem and related social issues. The purpose of the investigation was to determine whether students' argumentation in small group discussion can be categorized based on their ability to use evidence and identify its limitations. The study was guided by four questions: What are the characteristics of argumentation occurring in small

groups? What are the characteristics of students' understanding of the limitations of evidence? What interpretations do students make of the evidence they have in terms of the arguments they are making? and What group processes are undertaken by students when involved in interpreting evidence and understanding its limitations?

Characteristics of Argumentation

The findings on the characteristics of argument present in small group discussion indicated many of the groups demonstrated level 3 argumentation with the majority of these making claim(s) with reason(s) responded to by rebuttal and the others making competing claims with reason(s) given and qualifier(s). One group demonstrated level 4, seven groups demonstrated level 1, and two demonstrated level 0 argumentation characteristics. An earlier study in school B with one class of 21 students somewhat younger students, 12.6-13.8 versus this study's range of 13.6-15.6, found half of its students at level 3 with others at level 1 (Sunal & Sunal, 1999).

When examined by school, more student groups at school A demonstrated level 3 argumentation characteristics with one group demonstrating level 4. School B groups demonstrated lower levels, with fewer at level 3 and two groups at level 1. This difference may be partially explained by the higher mean age found at school A, 15.0 versus a mean of 14.2 at school B. However, some studies of reasoning and argumentation have found little change in the ability after the end of middle/junior high school (Perkins, 1985; Voss & Means, 1991; Kuhn, 1993). The conclusion is that maturation alone does not enhance reasoning and argumentation abilities. Hence, educational intervention such as that carried out by Shayer and Adey (1993) in fostering the development of specific formal reasoning skills appears to be necessary if reasoning and argumentation are to be developed. It is possible that the difference found between school A and B students may be due to factors other than age such as teacher instructional intervention or the experiential base of the students. The learning cycle used in the study's procedures may have fostered some development but student groups at both sites participated in the same learning cycle and Shayer and Adey's work indicates a long term intervention is needed for such development.

Since school B had a diverse set of classes with three different school system designations based on SAT-9 scores, grade point average, and special education assessments where applicable, the groups within each designation were also examined for similarities and differences. The class designations were: average to above average, gifted, and low performing. Groups in the average to above average classes mostly demonstrated level 3 characteristics with 35% demonstrating level 1. Groups in the class designated as gifted all demonstrated level 3 characteristics. In the class designated as low performing, just one group demonstrated level 3 characteristics while two demonstrated level 1 and two demonstrated level 0. The findings indicate that class designation is related to the level of argumentation characteristics demonstrated. There is less difference between the average to above average groups and the gifted-designated groups than there is between these groups and the groups designated as low performing. The average to above average groups had a larger range of students, some of whom are very close in test scores and grade point average to those students in the class designated as gifted. The findings suggest that the characteristics used in analyzing their argumentation reflect the level of development found in school B's students since the characteristics they demonstrate match to some extent their general designation.

Use of Evidence and Identification of its Limitations

The ability of the small groups to use evidence and identify its limitations during argumentation was examined using the instrument, Identification of Evidence and its Limitations in Argumentation. The findings are reported above in terms of the three subgroups of criteria found in the instrument.

Most of the student groups gathered evidence beyond that given them by their teacher with a few attempting exhaustive evidence gathering. Two groups at school B used only the evidence they obtained from their teacher. One of these groups was in the low performing designated class and one in an average to above average designated class. The findings using these criteria were very similar at both schools although school A's student groups always scored slightly higher in achieving each of the criteria used. This difference was small enough that the pattern for both schools should be considered similar. Within school B, some small differences are seen in use of the evidence base provided by the teacher: 100% of the groups in the class designated as gifted used it, while 92% of the average to above average groups and 67% of the low performing groups did so. With this criterion, the variance in numbers of groups with each designation should be considered. One group in the low performing and one group in the average to above average classes did not use the evidence base provided. Because there were more groups, 17 versus 6, the percentage is less effected by the performance of one group in the average to above average classes. The pattern across all of school B's groups is, therefore, a usage of the existing evidence base. There is much more divergence between the low performing groups and the other groups when gathering of evidence beyond that provided by the teacher is used as a criterion. Just one group of the low performing class gathers additional evidence while all of the other groups but one do so. Exhaustive evidence gathering was done by very few groups but these were found in the average to above average and gifted designated class groups and not at all in the low performing class groups. The general pattern that emerges differs the most when the second criterion is considered with students in the designated low performing class demonstrating much less effort at evidence gathering. Because of limitations of time, the small effort involving a few groups to do exhaustive evidence gathering could be expected. Such evidence gathering requires time to carry out, time to reflect on the available evidence, and time to decide how to acquire further evidence. Since this study was limited to a two-week period during which the learning cycle was carried out, there was little time for exhaustive evidence gathering. Student groups designated as low performing may have the most difficulty planning and carrying out a search for further evidence because of a lack of skills, a low experience base, or a greater need for additional time for reflection and organizing.

All student groups with the exception of one group in a class designated as average to above average were able to identify one limitation in the evidence. The two groups who did not examine all of the evidence base provided to them by their teacher examined parts of it and identified a limitation in those parts they examined. Most groups identified more than one limitation with a greater percentage doing so in school A than in school B. Within school B the percentage of groups identifying more than one limitation in the evidence within different designations followed a pattern similar to that in many of this study's findings: the group designated as gifted had the highest percentage of groups meeting this criterion, the average to above average group the second highest, and the low performing group the least. It should be noted, however, that half of the low performing groups identified more than one limitation and all of these groups identified one limitation. Therefore, it can be concluded that these student groups were able to identify limitations in evidence at a minimum level, and generally to a level

that was at more than a minimum. The last criterion in this subgroup was the context of the times. Since the topic investigated was dealt with nuclear fusion and its possibilities as a source of commercial energy, this criterion examined the recognition of factors in current society and the stage of scientific research and consideration of their influence on the questions being investigated by the groups. Student groups met this criterion less often than they met the other criteria in this subgroup with 55% of the groups doing so. All but two groups at school A met the criterion and half of school B's groups. Since students at school A were in classes one grade level higher than were students in school B this difference could be due to the additional social studies as well as additional science classes taken. The state course of study emphasized the use of primary source documents in social studies classes and considerations of the context of the times. So, additional participation is experiences with such a focus may have impacted school A's students. Further investigation of the potential influences of social studies on this aspect of argumentation is needed. Within school B, all of the groups in the class designated as gifted and half of the groups in the average to above average classes considered the context of the times. In the low performing designated class just one group met this criterion. Consideration of the context of the times requires information, the ability to weigh the influence of factors, and reflection (Lee, Dickinson, & Ashby, 1994). Research on students' understanding of history, however, indicates that the context of the times is difficult for middle school students (VanSledright, 1994). In this class the group that had not considered all of the evidence base provided by the teacher was able to identify only one limitation of the evidence and did not consider the context of the time. Also, the four groups who had not gathered evidence beyond that provided by the teacher also did not consider the context of the times. Because some of low performing designated groups either did not consider all of the evidence provided them and/or did not gather some evidence on their own, they had a weaker information base than did other groups. This weaker information base provides less opportunity for identifying potentially important factors relating to the questions under investigation and for weighing the differential influence of factors. The lower ability of some of the groups in this class to identify more than one limitation of the evidence may result from the weaker information base and also may indicate less ability at reflection and weighting factors. Further investigation is needed to establish patterns that may be occurring within a group on the criteria within this instrument.

While all the criteria in subgroups A and B deal with evidence, those in subgroup C consider the use of evidence in argumentation. All groups cited evidence in their argumentation and all but two groups, one designated as average to above average and one designated as low performing, linked two pieces of evidence in making an argument. The low performing group also had not examined the existing evidence base provided by the teacher. The evidence was organized into a meaningful structure during argumentation by 66% of the groups, with 60% in school A and 68% in school B doing so. As seen with some criteria discussed above, groups in the low performing designated class met this criterion less often than did other groups at school B, 33% (two of six groups) as compared to 71% of the average to above average and 83% of the gifted designated groups. The low performing designated group that had not considered all the evidence provided by the teacher and had not linked two pieces of evidence in making an argument nor met some of the other criteria, did not meet this criterion either. This criterion evaluated a difficult task. While groups may link evidence in a multitude of ways, the criterion was whether their final structure could be considered meaningful. Thus, pieces of evidence had to support each other and some recognition of limitations of the evidence was needed in order for the argument to be meaningful. Because earlier criteria in this instrument were not attained by a group, it was not possible for some groups to develop a meaningful structure for their

argumentation. Counter evidence was considered during argumentation by less than half of the groups (39%), but these were more often in school A (50%) than school B (21%). At school B, half of the gifted designated groups and 18% of the average to above average designated groups considered counter evidence. The consideration of counter evidence follows a pattern similar to that found with other criteria in this instrument: attainment of the criteria more often by the groups designated as gifted, somewhat less often by those designated as average to above average and least often by the low performing designated groups. A second, less frequent pattern is also discerned: school A's somewhat older student groups who are in a grade lever higher, are more successful than are school B's student groups.

The analysis of the findings resulting from the use of the instrument, Identification of Evidence and its Limitations in Argumentation, indicate that is does categorize the argumentation of students during their small group discussion based on their ability to use evidence and identify its limitations.

The last analysis of small group argumentation examined group usage of seven processes identified by Driver and Newton (1997) as facilitative of the involvement of all members. All but two groups, both in school B's low performing designated class, checked the information and evidence being presented. One of the groups in the low performing designated class was the same group that had not met many of the criteria examined when considering group identification of evidence and its limitations. The findings indicate that most of these student groups demonstrate the ability to check information and evidence being presented.

Within their groups, students questioned each other's reasons during argumentation with exception of two groups in the low performing designated class and one group in the average to above average class. Fewer, yet still most, of the groups (71%) checked the coordination of evidence and claims made during discussion. A higher percentage of these groups were in school A (80%) than in school B (68%). Use of this group process followed a pattern somewhat different from that found in other parts of the study: the average to above average designated groups (76%) slightly more often checked the coordination of evidence and claims than did the gifted designated groups (67%), but the difference was small. However, the low performing groups (33%) much less frequently carried out such checking. Building on each other's arguments was found in fewer groups (53%) with little difference between the two school sites. As with the coordination of evidence, average to above average designated groups (59%) somewhat more often build on each other's arguments than did groups designated as gifted (50%), but the difference was small. However, just one group in the low performing designated class did so. This group process was used by half of the groups indicating these student groups had some competency in the process. The findings indicate a need to foster the usage of this process in the groups not demonstrating it. Of all of the group processes discussed so far, this is most in need of development by students and the findings suggest there is potential for such development.

Two other group processes that foster the participation of all in discussion were analyzed. It was found that most of the groups (84%) monitored to insure each group member was involved. Monitoring occurred most in the gifted designated class (83%) and least in the low performing designated class (67%). While there was a difference between these two designations, 76% of the average to above average designated groups engaged in monitoring, demonstrating little difference from either of the other two sets of groups. Monitoring to insure each members involvement is a group process skill found widely among these groups. Efforts at further facilitation of this process in groups not demonstrating it appear to have potential for success since it has already developed among the majority of the groups. A second group process skill

fostering participation is encouragement of the presentation of different ideas. This process was found in the discussion of fewer groups (68%) than was monitoring. Little difference was found between student groups in schools A and B in the use of this process. Little difference was also found in usage between the school B groups designated as average to above average (71%) and gifted (67%). However, the low performing designated groups utilized this process to a lower extent, with usage found in 50% of these groups.

A final process examined was distinguishing between scientific claims and those based on other types of knowledge. This process was least used (39%) by the student groups. More of school A's (50%) groups used it than did school B's groups (36%). This difference may be due to maturity, additional experience in science classes, or other factors not investigated in this study. In school B a higher percentage of the average to above average designated groups (47%) used the process than did those groups designated as gifted or low performing, where it was found used by 17% of their designated groups. Although all of the group processes have relevance in consideration of students use of evidence in argumentation and recognition of its limitations, this group process can be considered to most strongly impact such argumentation. Evidence plays a central role in scientific claims. The limitations of evidence also play a role. Evidence with abundant limitations has restricted applicability. As the applicability is increasingly restricted, the use of the evidence in argumentation is also restricted. Distinguishing scientific claims requires students to have an appropriate information base, to have developed a range of skills in processing such information, and to reflect on the claims made. This is a complex set of requirements. This complexity appears to have impacted student groups usage of this process.

The findings of this study indicate that these student groups had developed some of the characteristics of argumentation and group process skills used in building and arguing an evidentiary case which recognizes the limitations of the evidence. However, these students and their student groups need guidance if they are to further develop their use of evidence and recognition of its limitations in argumentation. The findings further indicate that it was possible to categorize students' argumentation in small group discussion based on their ability to use evidence and identify its limitations.

References

American Association for the Advancement of Science, (1993). Benchmarks for scientific literacy. New York: Oxford University Press.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). Young people's images of science. Bristol, PA: Open University Press.

Driver, R. & Newton, P. (1997). Establishing the norms of scientific argumentation in classrooms. Paper presented at the European Science Education Research Association Conference, Rome.

Greene, S. (1994). Students as authors in the study of history. In G. Leinhardt, I. Beck, and C. Stainton (Eds.), Teaching and learning in history. Hillsdale, NJ: Lawrence Erlbaum.

Kuhn, D. (1993). Science as argument: implications for teaching and learning scientific thinking. Science Education 77(3), 319-337.

Kuhn, T. S. (1962). The structure of scientific revolutions (2nd ed.). Chicago: University of Chicago Press.

Lee, P., Dickinson, A., & Ashby, R. (1994). Researching children's ideas about history. Paper presented at Madrid conference, Cognitive and Instructional Processes in History, Madrid, Spain.

Leinhardt, G., Beck, I., & Stainton, C. (Eds.) (1994). Teaching and learning history. Hillsdale, NJ: Lawrence Erlbaum.

Leinhardt, G., Stainton, C., Virji, S., & Odoroff, E. (1992). Learning to reason in history: mindlessness to mindfulness. Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

Levstik, L. & Pappas, C. (1987). Exploring the development of historical understanding. Journal of Research and Development in Education 21, 1-15.

National Research Council, (1995). National science education standards, Washington, D.C.: National Research Council.

Norris, S. P. & Phillips, L. M. (1994). Interpreting pragmatic meaning when reading popular reports of science. Journal of Research in Science Teaching 31(9), 947-967.

Osbourne, J. (November, 1997). The structure and organisation of the science curriculum. Paper presented at the Science Education Seminar, King's College: London.

Perkins, D. (1985). Postprimary education has little impact on informal reasoning. Journal of Educational Psychology 77, 562-571.

Popper, K. R. (1934). Logik der forschung. Published in English (1959) as The logic of scientific discovery. London: Hutchinson.

Salter, D. (1997). Evaluation of evidence in historical text and recognition of historical authorial intention. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Shayer, M. & Adey, P. (1990). Accelerating the development of formal thinking in middle and high school students. Journal of Research in Science Teaching 27(3), 267-285.

Shayer, M. & Adey, P. (1993). Accelerating the development of formal thinking in middle and high school students IV: three years after a two-year intervention. Journal of Research in Science Teaching 30(4), 351-366.

Sunal, C. & Sunal, D. (1999). Scientific reasoning: exploring characteristics of middle school students' argumentation. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Tolumin, S. (1958). The uses of argument. Cambridge, UK: Cambridge University Press.

Voss, J. & Means, M. (1991). Learning to reason via instruction in argumentation. Learning and Instruction, 1, 337-350.

Wineburg, S. & Fournier, J. (1994). Finding home in a foreign country: the nature of contextualized thinking in history. In M. Carretero and J. Voss (Eds.), Cognitive and instructional processes in history and the social sciences (pp. 285-308). Hillsdale, NJ: Lawrence Erlbaum.

Yerrick, R. K. (2000). Lower track science students' argumentation and open inquiry instruction. Journal of Research in Science Teaching 37(8), 807-838.

Table 1
Class Data

School Site No. of Sites = 2	Class No. of Classes = 7	Class Size n=175	Class Composition females= 86 males=89	Student Age Range and Mean Age Total age range=13.6-15.6 Total group mean age=14.4
A	one	25	13 females/12 males	14.7-15.6 mean = 14.9
	two	25	11 females/14 males	14.7-15.3 mean = 15.1
B	three	25	13 females/12 males	13.7-14.1 mean = 13.9
	four	26	14 females/12 males	13.6-14.5 mean =14.2
	five	24	11 females/13 males	14.1-15.1 mean = 14.7
	six	25	11 females/14 males	13.8-14.4 mean = 14.2
	seven	25	13 females/12 males	13.7-14.1 mean = 13.9

Table 2
Characteristics of Small Group Argument

Characteristic of Argument

Group /Class	1.Single claim with no reasons given-level 0	2. Competing claims with no reasons given-level 0	3. Single claim with reason(s) given-level 1	4. Competing claims with reasons(s) given and qualifier(s)-level3	5. Claim(s) with reason(s) responded to by rebuttal-level 3	6. Making judgment(s) integrating different arguments-level 4
1/one					x	
2/one				x		
3/one					x	
4/one				x		
5/one					x	
6/two				x		
7/two				x		
8/two					x	
9/two					x	
10/two						x
11/three					x	
12/three				x		
13/three					x	
14/three					x	
15/three					x	
16/three				x		
17/four			x			
18/four					x	
19/four					x	
20/four			x			
21/four				x		
22/four					x	
23/five		x				
24/five	x					
25/five			x			
26/five			x			
27/five				x		

28/six					x	
29/six					x	
30/six				x		
31/six			x			
32/six				x		
33/seven					x	
34/seven			x			
35/seven				x		
36/seven					x	
37/seven			x			
38/seven			x			
Totals	1	1	8	11	16	1

Table 3
Summary by Class, Class Designation, and School Site of Characteristics of Argument

Class/School Class Designation	Characteristic					
Class	1. Single claim with no reasons given-level 0 (number and percentage of total groups)	2. Competing claims with no reasons given-level 0 (number and percentage of total groups)	3. Single claim with reason(s) given-level 1 (number and percentage of total groups)	4. Competing claims with reason(s) given and qualifier(s)-level 3 (number and percentage of total groups)	5. Claim(s) with reason(s) given and qualifier(s)-level 3 (number and percentage of total groups)	6. Making judgment(s) integrating different arguments-level 4 (number and percentage of total groups)
One (5 groups)				2 (5%)	3 (8%)	
Two (5 groups)				2 (5%)	2 (5%)	1 (3%)
Three (6 groups)				2 (5%)	4 (11%)	
Four (6 groups)			2 (5%)	1 (3%)	3 (8%)	
Five (5 groups)	1 (3%)	1 (3%)	2 (5%)	1 (3%)		
Six (5 groups)			1 (3%)	2 (5%)	3 (8%)	
Seven (6 groups)			3 (5%)	1 (3%)	2 (5%)	
Subtotals of no. of groups by characteristic	1 (3%)	1 (3%)	78(21%)	11 (29%)	16 (42%)	1 (3%)

School Mean and Percentage of Groups in School						
A				4 (40%)	5 (50%)	1 (10%)
B	1 (4%)	1 (4%)	8 (29%)	7 (25%)	12 (50%)	
Class designation within School B						
Average to above average (3 classes –#'s 4,6,7/17 groups)			6 (35%)	4 (24%)	8 (47%)	
Gifted class (1 class-#3/6 groups)				2 (33%)	4 (67%)	
Low-performing class (1 class-#5/6 groups)	1 (17%)	1 (17%)	2 (33%)	1 (17%)		

Note: percentages are rounded off

Table 4
Summary of Whole Group Characteristics of Argument

Characteristics	Number and percentage of total students
1. Single claim with no reasons given-level 0	1 (3%)
2. Competing claims with no reasons given-level 0	1 (3%)
3. Single claim with reasons(s) given-level 1	7 (18%)
4. Competing claims with reasons(s) given and qualifier(s)-level 3	11 (29%)
5. Claim(s) with reason(s) given and qualifier(s)-level 3	16 (42%)
6. Making judgment(s) integrating different arguments-level 4	1 (3%)

Table 5
Small Groups' Ability to Use Evidence and Identify Its' Limitations, Criterion Subgroup A

Group/ Class	Criterion			
	1) Examination of an existing evidence base	2) Gathering of evidence beyond that provided by teacher	3) Exhaustive evidence gathering	Subtotal
1/one	x	x		2
2/one	x	x		2
3/one	x			1
4/one	x	x		2
5/one	x	x		2
6/two	x	x		2
7/two	x	x		2
8/two	x	x	x	3
9/two	x	x		2
10/two	x	x	x	3
11/three	x	x		2
12/three	x	x		2
13/three	x	x		2
14/three	x	x		2
15/three	x	x	x	3
16/three	x	x		2
17/four	x	x		2
18/four	x	x		2
19/four	x	x	x	3
20/four				0

21/four	x	x		2
22/four	x	x		2
23/five	x			1
24/five				0
25/five	x			1
26/five	x			1
27/five	x	x		2
28/six	x	x		2
29/six	x	x	x	3
30/six	x	x		2
31/six	x	x		2
32/six	x	x		2
33/seven	x	x	x	3
34/seven	x	x		2
35/seven	x	x		2
36/seven	x	x		2
37/seven	x	x		2
38/seven	x	x	x	3
Subtotals	36	32	7	76

Table 6

Summary by Class of Small Groups' Ability to Use Evidence and Identify Its' Limitations, Criterion Subgroup A

Class	Criterion			
	1) Examination of an existing evidence base	2) Gathering of evidence beyond that provided by teacher	3) Exhaustive evidence gathering	Subtotal and Percentage
one (5 groups)	5	4	0	9 (60%)
two (5 groups)	5	5	2	12 (80%)
three (6 groups)	6	6	1	13 (55%)
four (6 groups)	5	5	1	11 (61%)
five (5 groups)	4	1	0	5 (33%)
six (5 groups)	5	5	1	11 (73%)
seven (6 groups)	6	6	2	13 (72%)
Subtotals and percentages	36 (95%)	32 (84%)	7 (18%)	

Note: percentages are rounded off

Table 7
Small Groups' Ability to Use Evidence and Identify Its' Limitations, Criterion Subgroup B

Group/ Class	Criterion			
	4) Identification of one limitation of evidence available	5) Identification of more than one limitation of evidence available	6) Considering the context of the times	Subtotal
1/one	x	x	x	3
2/one	x	x		2
3/one	x		x	2
4/one	x	x	x	3
5/one	x	x	x	3
6/two	x	x	x	3
7/two	x	x	x	3
8/two	x	x	x	3
9/two	x	x		2
10/two	x	x	x	3
11/three	x	x	x	3
12/three	x	x		2
13/three	x		x	2
14/three	x	x	x	3
15/three	x	x		2
16/three	x	x	x	3
17/four	x			1
18/four	x	x	x	3
19/four	x	x	x	3
20/four	x		x	2
21/four				0
22/four	x	x	x	3
23/five	x			1
24/five	x			1
25/five	x	x		2
26/five	x	x		2
27/five	x	x	x	3
28/six	x	x	x	3
29/six	x	x		2
30/six	x			1
31/six	x	x	x	3
32/six	x	x	x	3
33/seven	x	x		2
34/seven	x	x	x	3

35/seven	x		x	2
36/seven	x			1
37/seven	x	x		2
38/seven	x	x		2
Subtotals	37	28	22	87

Table 8
Summary by Class of Small Groups' Ability to Use Evidence and Identify Its' Limitations, Criterion Subgroup B

Class	Criterion			
	4) Identification of one limitation of evidence available	5) Identification of more than one limitation of evidence available	6) Considering context of the times	Subtotals and Percentage
one (5 groups)	5	4	4	13 (86%)
two (5 groups)	5	5	4	13 (86%)
three (6 groups)	6	5	4	15 (83%)
four (6 groups)	5	3	4	17 (94%)
five (5 groups)	5	3	1	9 (60%)
six (5 groups)	5	4	3	12 (80%)
seven (6 groups)	6	4	2	12 (67%)
Subtotals and percentages	37 (97%)	28 (74%)	21 (55%)	

Note: percentages are rounded off

Table 9
Small Groups' Ability to Use Evidence and Identify Its' Limitations, Criterion Subgroup C

Class/Group	Criterion				
	7) Citation of evidence	8) Linkage of two pieces of evidence in an argument	9) Organization of evidence into a meaningful structure	10) Consideration of counter evidence	Subtotals
1/one	x	x	x	x	4
2/one	x	x			2
3/one	x	x			2
4/one	x	x	x		3
5/one	x	x	x	x	4
6/two	x	x	x	x	4
7/two	x	x	x		3
8/two	x	x			2
9/two	x	x	x	x	4
10/two	x	x		x	3
11/three	x	x	x	x	4
12/three	x	x			2
13/three	x	x	x	x	4
14/three	x	x	x		3
15/three	x	x	x		3
16/three	x	x	x	x	4
17/four	x	x			2
18/four	x	x	x	x	4
19/four	x	x	x		3
20/four	x	x			2
21/four	x				1
22/four	x	x	x		3
23/five	x	x			2
24/five	x				1
25/five	x	x	x		3
26/five	x	x			2
27/five	x	x	x		3
28/six	x	x	x		3
29/six	x	x			2
30/six	x	x			2
31/six	x	x	x		3
32/six	x	x	x		3
33/seven	x	x	x	x	4
34/seven	x	x	x		3
35/seven	x	x	x		3
36/seven	x	x	x		3

37/seven	x	x	x		3
38/seven	x	x	x	x	4
Subtotals	38	36	25	11	110

Table 10
Summary by Class of Small Groups' Ability to Use Evidence and Identify Its' Limitations, Criterion Subgroup C

Class	Criterion				
	7) Citation of evidence	8) Linkage of two pieces of evidence in an argument	9) Organization of evidence into a meaningful structure	10) Consideration of counter evidence	Subtotals and Percentage
one (5 groups)	5	5	3	2	15 (75%)
two (5 groups)	5	5	3	3	16 (80%)
three (6 groups)	6	6	5	3	20 (83%)
four (6 groups)	6	5	3	1	15 (63%)
five (5 groups)	5	4	2	0	11 (55%)
six (5 groups)	5	5	3	0	13 (54%)
seven (6 groups)	6	6	6	2	20 (83%)
Subtotals and percentages	38 (100%)	36 (95%)	25 (66%)	11 (29%)	

Note: percentages are rounded off

Table 11
Summary by Class, Class Designation, and School Site of Small Groups' Ability to Use Evidence and Identify Its' Limitations

Class	Criterion										
	1	2	3	4	5	6	7	8	9	10	Sub-total and percent-age of possible group score
one	5	4	0	5	4	4	5	5	3	2	37 (74%)
two	5	5	2	5	5	4	5	5	3	3	42 (84%)
three	6	6	1	6	5	4	6	6	5	3	48 (80%)
four	5	5	1	5	3	4	6	5	3	1	38 (63%)
five	4	1	0	5	3	1	5	4	2	0	25 (50%)
six	5	5	1	5	4	3	5	5	3	0	36 (72%)
seven	6	6	2	6	4	2	6	6	6	2	46 (77%)
	36 (95%)	32 (84%)	7 (18%)	37 (97%)	28 (74%)	22 (58%)	38 (100%)	36 (95%)	25 (66%)	11 (29%)	
School mean and percent age of groups in school											
A 10 groups	10 100%	9 90%	2 20%	10 100%	9 90%	8 80%	10 100%	10 100%	6 60%	5 50%	
B 28 groups	26 93%	23 82%	5 18%	27 96%	19 68%	14 50%	28 100%	26 93%	19 68%	6 21%	
Class desig-nation within School B											
Aver-age to	15 88%	16 94%	4 24%	16 94%	11 65%	9 53%	17 100%	16 94%	12 71%	3 18%	

above aver-age: 3 classes - #'s 4,6,7 /17 groups											
Gifted class: 1 class-#3/6 groups	6 100%	6 100%	1 17%	6 100%	5 83%	4 100%	6 100%	6 100%	5 83%	3 50%	
Low-perfor-ming class: 1 class-#5/6 groups	4 67%	1 17%	0 0%	6 100%	3 50%	1 17%	5 83%	4 67%	2 33%	0 0%	

Note: percentages are rounded off

Table 12
Summary of Whole Group Ability to Use Evidence and Identify its' Limitations

Criterion	Number and Percentage
Examination of an existing evidence base	36 (95%)
Gathering of evidence beyond that provided by teacher	32 (84%)
Exhaustive evidence gathering	7 (18%)
Identification of one limitation of evidence available	37 (97%)
Identification of more than one limitation of evidence available	28 (74%)
Consideration of the context of the times	21 (55%)
Citation of evidence	38 (100%)
Linkage of two pieces of evidence in an argument	36 (95%)
Organization of evidence into a meaningful structure	25 (66%)
Consideration of counter evidence	11 (29%)

Table 13
Group Processes Used

Group/Class Process	Check information and evidence	Question each other's reasons	Check coordination of evidence and claims	Build on each other's arguments	Monitor involvement of group members	Encourage presentation of different ideas	Distinguish between scientific claims and those based on other types of knowledge	Totals
1/one	x	x	x	x	x	x	x	7 100%
2/one	x	x	x	x	x			5 71%
3/one	x	x	x	x	x	x	x	7 100%
4/one	x	x	x		x	x		5 71%
5/one	x	x			x			3 43%
6/two	x	x	x	x	x	x	x	7 100%
7/two	x	x	x		x			4 57%
8/two	x	x			x	x		4 57%
9/two	x	x	x	x	x	x	x	7 100%
10/two	x	x	x	x	x	x	x	7 100%
11/three	x	x	x	x	x	x		6

								86%
12/three	x	x	x		x	x		5 71%
13/three	x	x			x			3 43%
14/three	x	x	x	x		x		5 71%
15/three	x	x	x	x	x	x	x	7 100%
16/three	x	x			x			3 43%
17/four	x	x	x		x	x		5 71%
18/four	x	x	x	x	x	x	x	7 100%
19/four	x	x	x	x	x	x	x	7 100%
20/four	x	x			x	x		4 57%
21/four	x	x	x	x				4 57%
22/four	x	x	x	x	x	x	x	7 100%
23/five					x			1 14%
24/five					x	x		2 29%
25/five	x							1 14%
26/five	x	x	x		x	x		5 71%
27/five	x	x	x	x	x	x	x	7 100%
28/six	x	x	x	x	x			5 71%
29/six	x	x	x	x		x	x	6 86%
30/six	x	x			x	x		4 57%
31/six	x				x			2 29%
32/six	x	x	x	x	x		x	6 86%
33/seven	x	x	x	x		x		5 71%
34/seven	x	x	x	x	x	x	x	7 100%
35/seven	x	x	x	x	x	x	x	7 100%
36/seven	x	x			x	x		4 57%
37/seven	x	x	x					3 43%
38/seven	x	x	x		x	x	x	6 86%
Totals	36 (95%)	34 (89%)	27 (71%)	20 (53%)	32 ((84%)	26 (68%)	15 (39%)	

Note: percentages are rounded off